

Исследование алгоритмов на базе нейросетевой архитектуры YOLO в задаче детектирования полипов на колоноскопических видеоданных

В.В. Хрящёв¹, Н.В. Котов¹, А.Л. Приоров¹

¹ Ярославский государственный университет им. П. Г. Демидова, ул. Советская, д. 14, г. Ярославль, 150003, Россия

Аннотация

Приведен анализ использования нейросетевых алгоритмов детектирования полипов толстой кишки, получаемых при колоноскопическом исследовании. Для обучения и тестирования алгоритмов глубокого машинного обучения использовалась база изображений Kvasir-SEG. В качестве нейросетевых архитектур использовались сети YOLOv6, YOLOR, YOLOv7, YOLOv7X, YOLOv8, предварительно обученные на базе изображений MS COCO. В связи с малым объемом изображений в базе Kvasir-SEG использовалась аугментация данных. В результате применения алгоритмов детектирования к тестовому набору эндоскопических изображений получены наивысшие значения метрик $AP@[0,25..0,75]$ – равной 98,4, и $AP@0,50$ – равной 98,6, для нейросетевого детектора на базе сети YOLOv8. По результатам сравнения предложенных алгоритмов с аналогами сеть YOLOv8 показала улучшение результата по метрике $AP@[0,25..0,75]$ на 5,9 в сравнении с аналогичной нейросетевой моделью HAOO_{YL}. Полученные результаты могут использоваться при разработке системы анализа видеопотока в эндоскопической системе, работающей в режиме реального времени при проведении колоноскопических исследований.

Ключевые слова

Эндоскопия, колоноскопические изображения, полипы кишечника, глубокое обучение, нейросетевая архитектура YOLO.

Study of Algorithms Based on YOLO Neural Network Architecture in the Problem of Polyp Detection on Colonoscopic Video Data

V.V. Khryashev¹, N.V. Kotov¹, A.L. Priorov¹

¹ Yaroslavl state university P. G. Demidov, st. Sovetskaya, 14, Yaroslavl, 150003, Russia

Abstract

An analysis was made of the use of neural network algorithms for the detection of colon polyps obtained during colonoscopy. The Kvasir-SEG image database was used to train and test deep machine learning algorithms. The networks YOLOv6, YOLOR, YOLOv7, YOLOv7X, YOLOv8 previously trained on the basis of MS COCO images were used as neural network architectures. Due to the small volume of images in the Kvasir-SEG database, data augmentation was used. As a result of applying the detection algorithms to the test set of endoscopic images, the highest values of the metrics $AP@[0,25..0,75]$ equal to 98,4 and $AP@0,50$ equal to 98,6 were obtained for the neural network detector based on the YOLOv8 network. According to the results of comparing the proposed algorithms with analogues, the YOLOv8 assessment showed an increase in the results for the $AP@[0.25..0.75]$ metric by 5.9 in searches with the previous model HAOO_{YL}. The results obtained can be used in the development of a video stream analysis system in an endoscopic system operating in real time during colonoscopy studies.

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

EMAIL: v.khryashev@uniyar.ac.ru (В.В. Хрящёв); n.kotov2@uniyar.ac.ru (Н.В. Котов); pri@uniyar.ac.ru (А.Л. Приоров)

ORCID: 0000-0003-2978-2459 (В.В. Хрящёв); 0000-0002-6207-473X (Н.В. Котов); 0000-0003-2984-340X (А.Л. Приоров)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Keywords

Endoscopy, images, polyps, deep learning, YOLO neural network architecture.

1. Введение

Полипы – аномальные разрастания ткани, выступающие над слизистой оболочкой. Они могут возникать в любом месте желудочно-кишечного тракта, но в основном обнаруживаются в колоректальной области и часто считаются предшественниками колоректального рака. В толстом кишечнике выделяют несколько типов классификации полипов: по количеству, по патоморфологическим данным, по особенностям прикрепления к стенке кишки [1, 2].

Размер колоректальных полипов может существенно различаться: от совсем небольших до весьма развитых (≥ 10 мм). Обычно с помощью эндоскопического оборудования можно обнаружить и удалить относительно крупные полипы. Небольшие полипы (≤ 5 мм) являются весьма сложными объектами для обнаружения [3, 4].

Колоноскопия – дорогостоящая и трудоемкая инвазивная процедура, требующая качественной подготовки кишечника, а также наполнения его воздухом во время обследования. Тем не менее, в среднем при подобных обследованиях не выявляется порядка 20% полипов. Таким образом, риск заболеть раком в значительной степени зависит от способности врача-эндоскописта обнаруживать полипы. Новые эндоскопические устройства и диагностические инструменты улучшили частоту обнаружения полипов [5-7], но проблема их пропуска остается по-прежнему актуальной [8, 9].

Задача автоматического обнаружения полипов на эндоскопических изображениях является также актуальной для специалистов, разрабатывающих системы анализа видеоизображений в медицине. В ранних научно-технических работах особое внимание уделялось алгоритмам машинного обучения, полагающимся на ручной подбор дескрипторов, основанных на таких признаках как цвет и текстура полипов [10]. Позднее методы, основанные на сверточных нейронных сетях, привлекли значительное внимание исследователей [11-14] и стали часто применяться при анализе общедоступных наборов медицинских данных [15-17].

В последнее время пристальное внимание уделяется одноэтапным сетям из-за их высокой производительности и точности. Эти сети объединяет единая конфигурация сети – магистральная часть, сеть пирамиды признаков и предсказательная часть. В рамках задачи обнаружения полипов в предыдущих работах были апробированы архитектуры сетей EfficientDet-D0, Faster R-CNN, RetinaNet50, RetinaNet101, YOLOv3+spp, YOLOv4, и ColonSegNet, а также модификации сети YOLOv5 [9]. В порядке перечисления этих архитектур эволюционировала точность предсказания и скорость обработки кадров за счет структурных модификаций и методов обработки изображений на этапе обучения нейронных сетей.

Таким образом, согласно данным, приведенным в научно-технической литературе, на сегодняшний день наиболее эффективным подходом к обнаружению полипов на эндоскопических изображениях является использование сверточных нейронных сетей. Обнаружение и локализация полипов обычно имеют решающее значение во время проведения диагностической колоноскопической процедуры.

Целью данной работы является автоматическое обнаружение полипов при проведении колоноскопических исследований. В дальнейшем рассматриваются алгоритмы обнаружения на базе современной нейросетевой архитектуры YOLO [18-20], прогнозирующие координаты рамки, внутри которой находится полип.

2. Архитектура сверточной нейронной сети YOLO

В рамках исследования выбраны последние версии детекторов семейства YOLO – YOLOv6, YOLOR, YOLOv7, YOLOv7X, YOLOv8. Каждая следующая архитектура содержит различные модификации – как архитектуры, так и методов процедуры обучения. Рассмотрим общую архитектуру такой нейронной сети на примере YOLOv7. Процесс обработки эндоскопического

изображения на основе алгоритма, использующего архитектуру YOLOv7 представлен на рисунке 1. На рисунке 2 представлены компоненты сети YOLOv7.

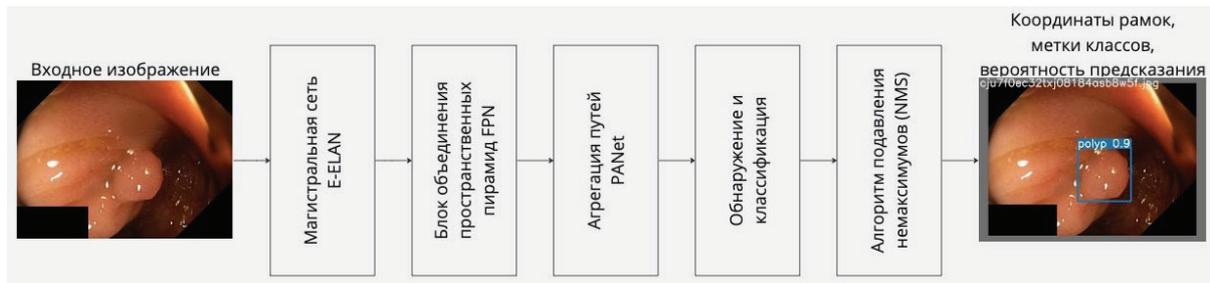
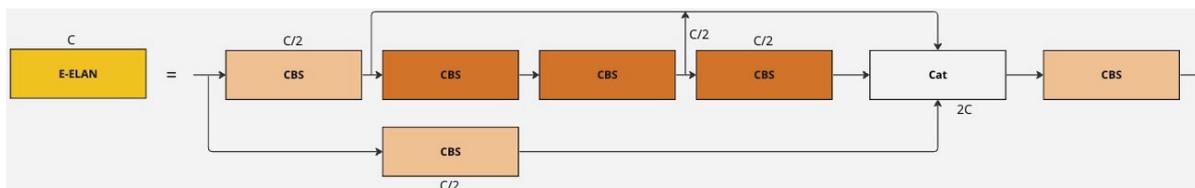
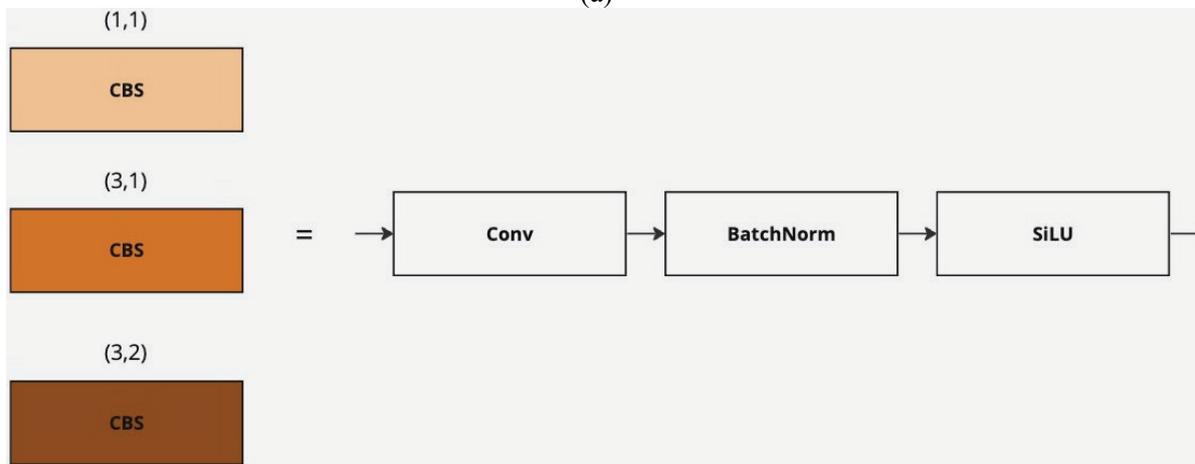


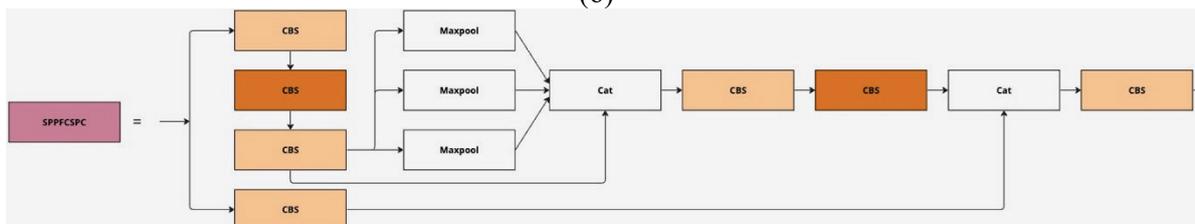
Рисунок 1 – Архитектура сети YOLOv7



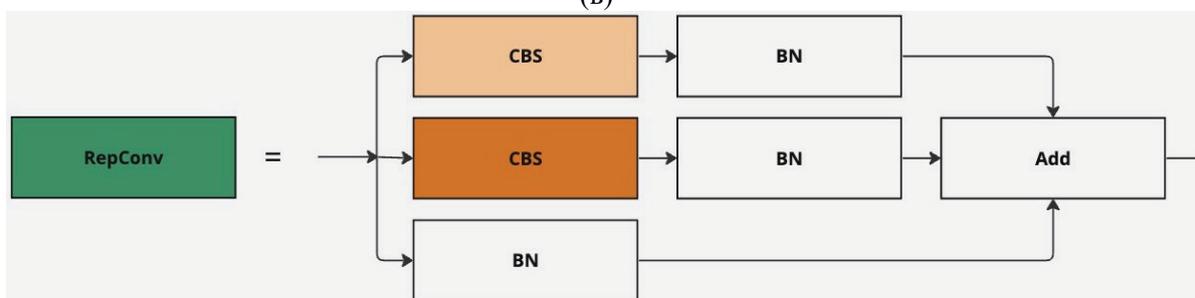
(a)



(б)



(в)



(г)

Рисунок 2 – Компоненты сети YOLOv7: а) блок-схема блока E-ELAN; б) блок-схема блока CBS; в) блок-схема блока SPPFCSPC; г) блок-схема блока RepConv

Нейронная сеть семейства YOLO состоит из трех частей – магистральная часть, часть объединения пространственных пирамид, часть предсказания.

В части сети, служащей для извлечения признаков объектов на входном изображении, используется сверточная сеть с использованием блоков E-ELAN (Extended Efficient Layer Aggregation Network). Блок E-ELAN контролирует кратчайший и самый длинный градиентный путь для лучшей сходимости модели, используя элементы расширения, перетасовки и слияния. В нем входные данные подаются непосредственно через блок свертки 1×1 , в то время как два других подключения к блоку получены путем свертки входных данных двумя блоками свертки 3×3 с тем же канальным множителем. Далее все функции объединены, и для изучения полной информации применяется свертка 1×1 .

К магистральной части добавляется часть объединения пространственных пирамид, выделяющая наиболее важные характерные особенности с разных уровней магистрали через сверточные слои блоков CBS (Convolution, BatchNormalization and SiLU) и SPPFCSPC (Spatial Pyramid Pooling Fast and Cross-Stage Partial Channel). Блок CBS является комбинацией сверточного слоя с разным размером ядра (1×1 , 3×1 и 3×2), слоя BatchNormalization (BN), и слоя активации SiLU (Sigmoid Linear Units). В конечных блоках пирамиды признаков свертки заменены блоками RepConv, которые необходимы для обновления параметров модели. Слои SPPFCSPC и RepConv являются комбинацией слоев CBS с разным размером ядра для слияния признаков на разных масштабах, при этом слой SPPFCSPC собирает признаки и изображения, а слой RepConv производит их слияние в части формирования предсказания сети.

Далее следуют слои обнаружения и классификации, дающие предсказание модели. Также используется механизм назначения меток, использующий главную и вспомогательные ветки предсказания и программный определитель меток. Тем самым, на этапе обучения модели используются признаки в промежуточных слоях, улучшающие итоговую предсказательную способность модели.

Для обучения указанных архитектур нейронных сетей требуются специализированные базы изображений с предварительной экспертной разметкой областей интереса. К сожалению, в отличие от собранных за последние два десятилетия гигантских по размеру баз изображений общего назначения (ImageNet, Pascal VOC, MS-COCO), в области эндоскопии подобные исследовательские проекты все еще находятся в начальной стадии развития. Рассмотрим существующую базу эндоскопических изображений более подробно.

3. База эндоскопических изображений полипов

Наборы размеченных экспертами изображений играют важную роль для построения алгоритмов анализа изображений методами глубокого машинного обучения. Рассмотрим одну из самых распространенных открытых баз эндоскопических изображений и видеоданных, полученных с помощью эндоскопа в режиме визуализации в белом свете, Kvasir-SEG [21]. Она содержит 1000 изображений толстой кишки с экспертной разметкой полипов для автоматического решения задач обнаружения и сегментации. База предоставляет возможность использовать этот набор как для решения задачи автоматической сегментации, так и для задачи автоматического обнаружения полипов. Таким образом, для исследований в данной работе использовался именно этот набор тестовых изображений.

На рисунке 3 показаны изображения, бинарные маски и ограничивающие рамки для полипов из набора Kvasir-SEG. Данный набор содержит изображения полипов, полученных с помощью эндоскопической системы Olympus, а также соответствующие им маски и информацию об ограничивающих рамках. Изображения и их маски могут использоваться для решения задачи сегментации полипов, а ограничивающие рамки – для их обнаружения.

Общее количество полипов в наборе данных Kvasir-SEG составляет 1072. Разрешение изображений варьируется от 332×487 до 1920×1072 пикселей. Набор данных включает изображения 700 «больших полипов» ($> 160 \times 160$ пикселей), 323 полипов «среднего размера» ($> 64 \times 64$ пикселей и $\leq 160 \times 160$ пикселей) и 48 «маленьких полипов» ($\leq 64 \times 64$ пикселей). Следует отметить, что такое число обучающих примеров на современном этапе развития нейросетевых технологий обнаружения объектов является недостаточным, так как типичное число таких

изображений обычно составляет десятки и даже сотни тысяч штук. С учетом этого необходимым этапом в построении алгоритма обнаружения полипов является этап аугментации (размножения) данных, который должен производиться с учетом специфики эндоскопических изображений [8-9].

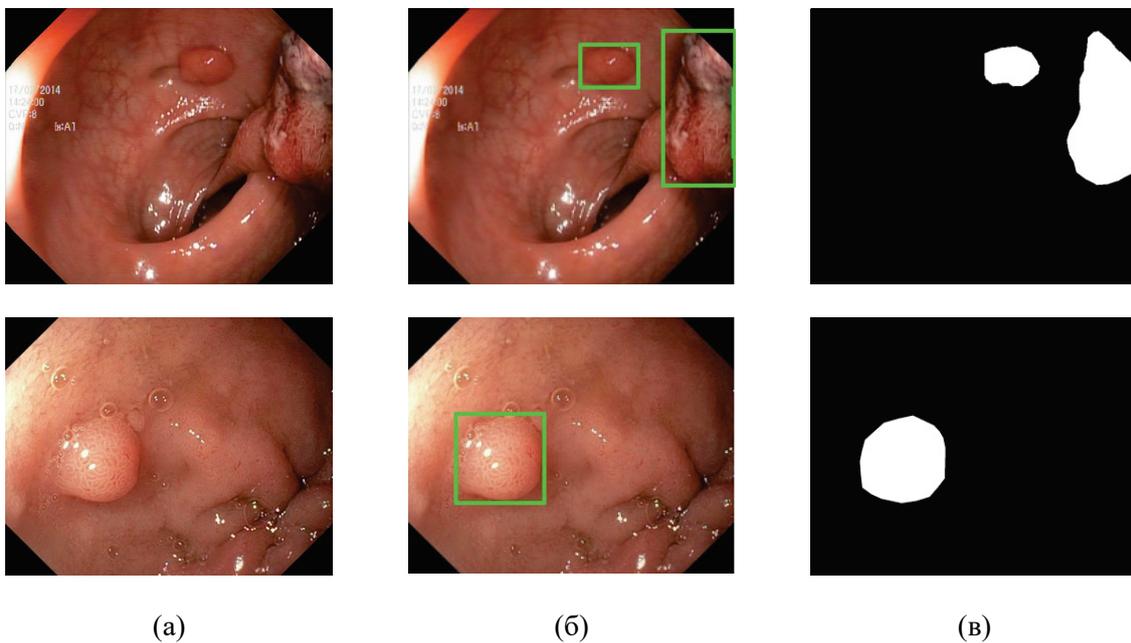


Рисунок 3 – Примеры изображений из набора данных Kvasir-SEG: а) исходные изображения; б) изображения с ограничивающими рамками; в) аннотированные бинарные маски

Общее количество полипов в наборе данных Kvasir-SEG составляет 1072. Разрешение изображений варьируется от 332×487 до 1920×1072 пикселей. Набор данных включает изображения 700 «больших полипов» ($> 160 \times 160$ пикселей), 323 полипов «среднего размера» ($> 64 \times 64$ пикселей и $\leq 160 \times 160$ пикселей) и 48 «маленьких полипов» ($\leq 64 \times 64$ пикселей). Следует отметить, что такое число обучающих примеров на современном этапе развития нейросетевых технологий обнаружения объектов является недостаточным, так как типичное число таких изображений обычно составляет десятки и даже сотни тысяч штук. С учетом этого необходимым этапом в построении алгоритма обнаружения полипов является этап аугментации (размножения) данных, который должен производиться с учетом специфики эндоскопических изображений [8-9].

4. Тестирование нейросетевых алгоритмов обнаружения полипов

Для алгоритмов на основе архитектур YOLO использовались следующие параметры обучения: размер батча (последовательности изображений) составлял 16 изображений, в качестве алгоритма оптимизации использовался алгоритм Adam (Adaptive Moment Estimation). Он сочетает в себе как идею накопления движения, так и идею более слабого обновления весов для типичных признаков. Начальная скорость обучения 10^{-3} , общее число эпох обучения – 250. Другие параметры алгоритма выбирались следующим образом: $\beta_1 = 0,9$; $\beta_2 = 0,999$; $\epsilon = 10^{-8}$.

Во время обучения использовались функции потерь BCE с Logits Loss для сетей YOLOv6, YOLOv7 и YOLOR, а также BBox (IoU) – для сети YOLOv8.

Для сравнения с аналогичными алгоритмами, приведенными в [9, 22], база эндоскопических изображений Kvasir-SEG разделялась на тестовый, валидационный и обучающий набор: 120 – тестовый и 880 – обучающий. Также, для обучения и тестирования сетей координаты ограничивающих рамок в файлах разметки базы пересчитывались из формата Pascal-VOC в формат YOLO.

Для оценивания качества работы алгоритмов в [22] предложено использовать метрики средней точности, рассчитанные при значениях порога IoU 0,25, 0,50 и 0,75. Далее данные метрики обозначены как $AP@0,25$, $AP@0,50$ и $AP@0,75$ соответственно. Кроме того, для комплексной оценки по множеству порогов IoU предложена метрика $AP@[0,25..0,75]$, которая рассчитывается как среднее 11-ти значений AP , рассчитанных при 11 различных порогах IoU в пределах от 0,25 до 0,75 с шагом 0,05 [9].

В таблице 1 приведены результаты тестирования алгоритмов на всех архитектурах на базе Kvasir-SEG.

Таблица 1 – Тестирование алгоритмов детектирования

Метрика	$AP@[0,25..0,75]$	$AP@0,25$	$AP@0,5$	$AP@0,75$
Известные аналоги из [22]				
EfficientDet-D0	47,6	68,5	50,5	22,8
Faster R-CNN	78,7	89,5	84,2	56,6
RetinaNet50	87,0	93,9	90,9	69,7
RetinaNet101	87,4	94,8	90,9	71,3
YOLOv3+spp	81,0	88,6	85,3	75,9
YOLOv4	85,1	91,2	82,3	75,9
ColonSegNet	80,0	90,0	81,7	67,1
Предложенные алгоритмы из [9]				
HAOO _{SV512}	90,2	96,6	93,8	74,9
HAOO _{YS}	88,6	94,0	90,8	73,8
HAOO _{YM}	91,8	95,9	93,3	82,8
HAOO _{YL}	92,5	96,7	93,9	85,1
Алгоритмы, исследованные в настоящей работе				
YOLOv6	89,2	93	90	84,1
YOLOv7	87,8	93,2	88,2	79,3
YOLOv7X	80,6	89,5	80,6	66,4
YOLOv7X	80,9	88,7	80,9	70
YOLOv8	98,4	98,7	98,6	97,6

На рисунке 4 представлены графики зависимости точности от полноты на тестовых изображениях для всех алгоритмов детектирования.

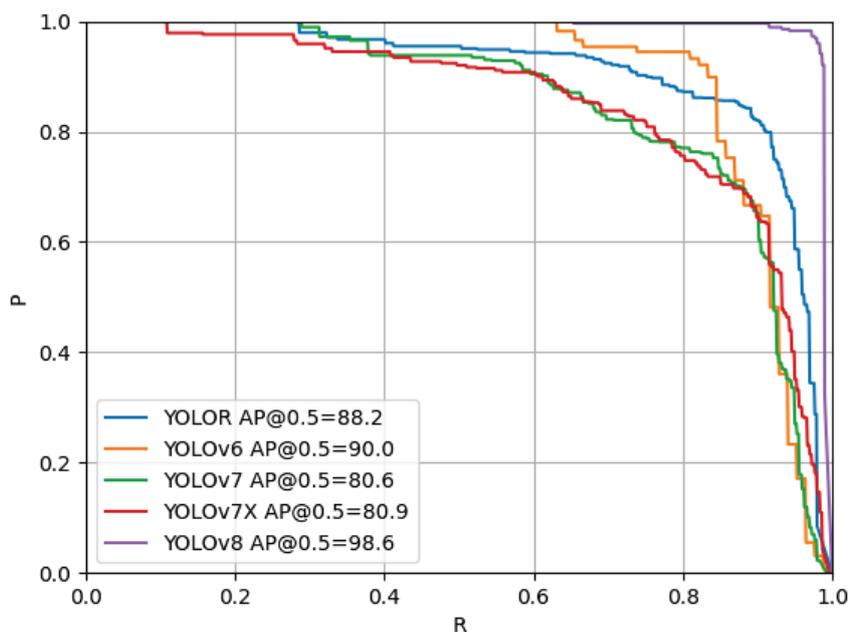


Рисунок 4 – Зависимости точности от полноты для представленных нейросетевых алгоритмов при тестировании на базе Kvasir-SEG

Среди представленных алгоритмов детектирования лучшие результаты показал YOLOv8, согласно результатам из таблицы 1. Значение $AP@0,50$ достигает для него 98,6, а усреднённая по диапазону порогов метрика $AP@[0,25..0,75]$ – 98,4. Это говорит о высокой способности детектирования объектов в широком диапазоне порогов детектирования.

Стоит отметить, что тестирование сетей YOLOv7, YOLOv7X и YOLOR показало значение метрик $AP@0,50$ меньше, чем для YOLOv6, несмотря на обратные показатели по данной метрике на больших эталонных базах. Это связано с большим числом параметров сети относительно количества обучающих примеров в базе Kvasir-SEG, несмотря на применение методов аугментации во время процесса обучения.

Таким образом, установлено, что алгоритм YOLOv8 достигает наилучшего результата по метрикам средней точности среди всех рассматриваемых алгоритмов обнаружения полипов. Его преимущество особенно существенно по метрике $AP@0,75$.

5. Заключение

Результаты проведенных исследований в части разработки алгоритмов обнаружения полипов на изображениях колоноскопических исследований позволяют сделать следующие основные выводы.

- На сегодняшний день наблюдается существенное отставание размера баз эндоскопических изображений от современных алгоритмов и методов глубокого машинного обучения.
- В результате тестирования предложенных алгоритмов обнаружения патологических объектов на изображениях с использованием открытого набора изображений Kvasir-SEG наилучший результат получен для нейросетевого алгоритма YOLOv8, для которого значение метрик $AP@[0,25..0,75]$ равно 98,4 и $AP@0,50$ – 98,6 соответственно.
- По результатам сравнения алгоритмов, приведенных в настоящем исследовании, и других исследований в области детектирования полипов на колоноскопических снимках можно сделать следующий вывод – предложенный алгоритм YOLOv8 значительно превосходит приведенные в более ранних исследованиях аналоги по метрике AP с всеми порогами. Относительно старшей модели HAOO_{YL} сеть YOLOv8 имеет преимущество на 5,9 по метрике $AP@[0,25..0,75]$. С другой стороны, более младшие модели не показали превосходства по данным метрикам.
- Исследования, проведенные в работе, послужат основой для построения на основе разработанного нейросетевого алгоритма обнаружения полипов модуля анализа видеопотока в реальной эндоскопической системе обнаружения полипов.

6. Список источников

- [1] Палевская С.А., Короткевич А.Г. Эндоскопия желудочно-кишечного тракта. М.: ГЭОТАР-Медиа, 2020. 752 с.
- [2] Куваев Р.О., Кашин С.В. Современное эндоскопическое исследование желудка с использованием методик узкоспектральной и увеличительной эндоскопии: техника проведения и алгоритмы диагностики // Доказательная гастроэнтерология. 2016. Т. 2, № 5. С. 3–13.
- [3] Lee J. Resection of diminutive and small colorectal polyps: what is the optimal technique? Clinical endoscopy. 2016. vol. 49(4). pp. 355.
- [4] Convolutional neural network for early detection of gastric cancer by endoscopic video analysis / A. Lebedev [et al.] // Proc. SPIE 11433. Twelfth International Conference on Machine Vision (ICMV 2019).
- [5] Батухтин Д.М., Пеганова Е.В., Митракова Н.Н. Анализ узкоспектральных эндоскопических изображений на внутренней поверхности пищевода // Вестник Поволжского

- государственного технологического университета. Серия: радиотехнические и инфокоммуникационные системы. 2014. № 4 (23). С. 45–57.
- [6] Коваленко Д.А., Гнатюк В.С. Ассоциация сцен в эндоскопических видео // GraphiCon 2017: Обработка и анализ биомедицинских изображений. Пермь. 2017. С. 269–274.
- [7] How to improve the adenoma detection rate in colorectal cancer screening? Clinical factors and technological advancements / M. Matyja [et al.] // Archives of medical science: AMS. 2019. vol. 15(2). pp. 424.
- [8] Разработка алгоритма детектирования полипов на эндоскопических изображениях с использованием сверточных нейронных сетей / А.А. Лебедев [и др.] // Цифровая обработка сигналов. 2021. № 2. С. 55-60.
- [9] Лебедев А.А. Исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях в медицинских системах прикладного телевидения: дис. к-та техн. наук: 2.2.13 / ЯрГУ. Ярославль, 2022. 131 с.
- [10] Computer-aided tumor detection in endoscopic video using color wavelet features / S.A. Karkanis [et al.] // IEEE Trans. on Information Technology in Biomedicine. 2003. vol. 7(3). pp. 141–152.
- [11] Николенко С.И., Кадурын А.А., Архангельская Е.О. Глубокое обучение. Питер: СПб., 2018. 480 с.
- [12] Convolutional neural networks for medical image analysis: Full training or fine tuning? / N.Tajbakhsh [et al.] // IEEE transactions on medical imaging. 2016. vol. 35(5). pp. 1299–1312.
- [13] Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning / H.-C. Shin [et al.] // IEEE Trans. on Medical Imaging. 2016. vol. 35(5). pp. 1285–1298.
- [14] Polyp-alert: Near real-time feedback during colonoscopy / Y. Wang [et al.] // Computer Methods and Programs in Biomedicine. 2015. vol. 120(3). pp. 164-179.
- [15] Automatic Colon Polyp Detection Using Region Based Deep CNN and Post Learning Approaches / Y. Shin [et al.] // IEEE Access. 2018. vol. 6. pp. 40950–40962.
- [16] Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets / J. Lee [et al.] // Scientific Reports. 2020. vol. 10(1). pp. 1–9.
- [17] Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy / M. Yamada [et al.] // Scientific reports. 2019. vol. 9(1). pp. 1–9.
- [18] YOLOv6: A single-stage object detection framework for industrial applications / Li C. [et al.] [Электронный ресурс] // arXiv preprint arXiv:2209.02976. 2022. (дата обращения 09.07.2023).
- [19] Wang C.-Y., Bochkovskiy A., and Liao H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [Электронный ресурс] // arXiv preprint arXiv:2207.02696. 2022. (дата обращения 09.07.2023).
- [20] G. Jocher, A. Chaurasia, and J. Qiu, YOLO by Ultralytics [Электронный ресурс] URL: <https://github.com/ultralytics/ultralytics>, 2023. (дата обращения 09.07.2023).
- [21] Kvasir-seg: A Segmented Polyp Dataset / D. Jha [et al.] // Proc. of International Conference on Multimedia Modeling. 2020. pp. 451–462.
- [22] Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning / D. Jha [et al.] // Computer Vision and Pattern Recognition. 2021. pp. 40496–40510.